



REPORT SEMINAR: FACE RECOGNITION USING 3D IMAGES PATTERN RECOGNITION

Instructors: Lê Hoàng Thái
Nguyễn Ngọc Thảo

Students:

- | | |
|-----------------------|---------|
| 1. Huỳnh Lê Minh Nhật | 1712632 |
| 2. Châu Thiên Thanh | 1712153 |
| 3. Phạm Minh Thắng | 1712759 |



Mục lục

I. ĐẶT VẤN ĐỀ:	3
II. GIỚI THIỆU TỔNG QUAN:	3
III. PHƯƠNG PHÁP:	4
1. 3D Face Recognition: UR3D	4
1.1. Tiền xử lý dữ liệu:	5
1.2 Annotated Face Model (AFM)	5
1.3. Registration (Pose correction - căn chỉnh góc mặt):	6
1.4. Deformable Model Fitting	7
1.5. Phân tích ảnh Geometry:	7
1.6. Các độ đo cho bước xác thực:	8
2. 3D Face Recognition for Partial Scans: UR3D-PS:	9
2.1 3D Landmark Detection:	9
2.2 Symmetric Deformable Model Fitting:	11
3. 3D-aided 2D Face Recognition: UR2D:	11
3.1 3D + 2D Enrollment:	11
3.2 2D Authentication:	12
3.3 Skin Reflectance Model :	13
IV. THỰC NGHIỆM:	13
1. 3D Face Recognition: UR3D	13
2. 3D Face Recognition for Parital Scans: UR3D-PS	14
V. DEMO:	16
VI. STATE OF THE ART:	17
1.. Method:	17
1.1 FR3DNet:	17
1.2 K3DM:	18
2. Performance:	19
VII. REFERENCES	20

I. Đặt vấn đề:

Mặc dù những nghiên cứu trong lĩnh vực nhận dạng khuôn mặt đã kéo dài hơn ba thập kỷ, nhưng đa phần chỉ có các hệ thống nhận dạng 2D là được phát triển mạnh cho đến nay. Tuy nhiên, nhận dạng khuôn mặt 2D vẫn còn hạn chế về độ chính xác, với các lý do chính sau đây:

- Sự khác biệt về góc mặt giữa tập xác thực và tập khảo sát
- Sự thay đổi về ánh sáng
- Sự đa dạng về biểu cảm gương mặt

Hiện nay, đi cùng với sự phát triển của công nghệ quét 3D thì giải pháp nhận dạng khuôn mặt sử dụng ảnh 3D cũng được đề xuất nhằm giải quyết những vấn đề của nhận dạng khuôn mặt 2D. Tuy nhiên bản thân giải pháp 3D cũng gặp phải những thách thức cản trở việc áp dụng vào thực tế, mà điểm mấu chốt nằm ở chi phí và quá trình quét 3D. Những giải pháp được giới thiệu sau đây sẽ góp phần khắc phục những vấn đề trên.

II. Giới thiệu tổng quan:

Đầu tiên, phải nói đến hệ thống nhận dạng khuôn mặt 3D hoàn toàn tự động UR3D với ý tưởng sử dụng Annotated Face Model (AFM) để mô tả dữ liệu về khuôn mặt. AFM được dùng quét khuôn mặt bằng cách sử dụng deformable model framework, xác định các chi tiết của khuôn mặt dưới dạng hình học 3D rồi đem đi biểu diễn dưới dạng 2D bằng cách sử dụng các tham số của model. Cách tiếp cận này linh hoạt hơn do sử dụng những mô hình hình học của mặt người, hiểu rõ hơn các vấn đề về nhận diện khuôn mặt và không cần phải huấn luyện

Thứ hai, chúng ta giải quyết vấn đề biến thể về góc mặt trong nhận dạng khuôn mặt 3D. Trong thực tế dữ liệu bị thiết thường xuyên xảy ra do việc quét 3D một phần khuôn mặt (thường là quét phần trước khuôn mặt). Điều này được giải quyết bằng cách xác định landmarks trong mỗi lần quét và sau đó registration với AFM, độc lập với góc mặt ban đầu của khuôn mặt. Đối với quét không trực diện, dữ liệu bị thiếu có thể bù đắp bằng cách khai thác tính đối xứng của khuôn mặt hoặc cố gắng nhận dạng dựa trên cấu hình khuôn mặt (bộ dữ liệu khuôn mặt có sẵn ở các góc mặt khác nhau đem đi so sánh với dữ liệu đầu vào).

Cuối cùng là giải quyết vấn đề về chi phí của máy quét 3D bằng phương pháp lai (kết hợp giữa máy quét 3D và máy ảnh 2D). Sử dụng dữ liệu 2D + 3D (hình dạng là 3D nhưng cấu trúc là 2D) để xây dựng các mô hình 3D. Phương pháp này đã cải thiện đáng kể độ chính xác so với hệ thống nhận diện khuôn mặt 2D trong những trường hợp khó.

III. Phương pháp:

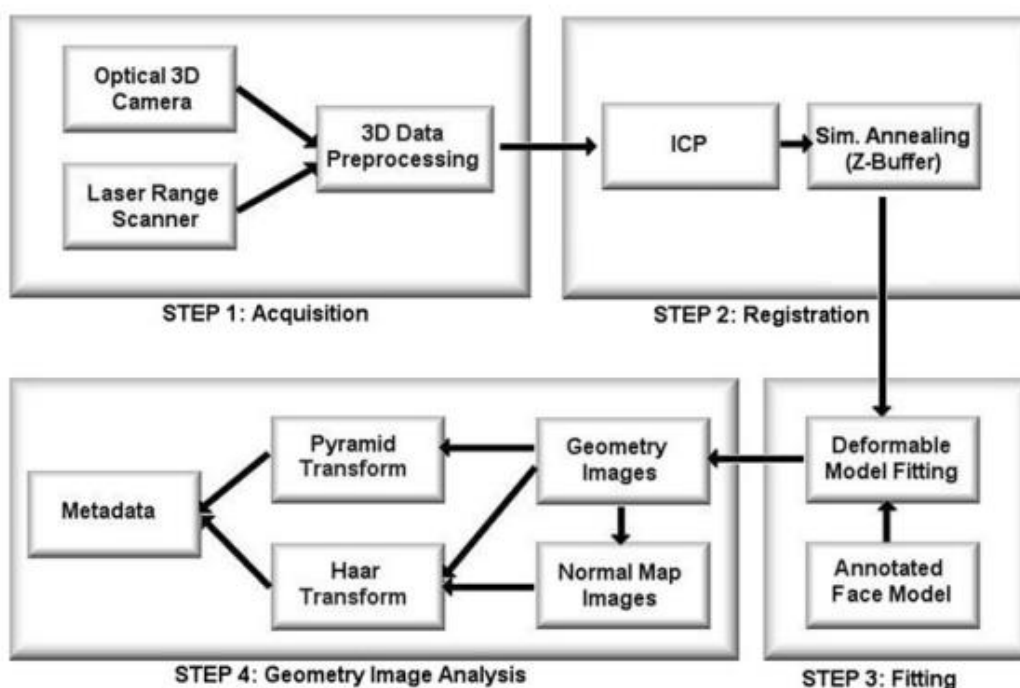
1. 3D Face Recognition: UR3D

Ý tưởng cơ bản của hệ thống này là mô tả dữ liệu khuôn mặt bởi mô hình Annotated Face Model (AFM).

Quy trình nhận dạng khuôn mặt 3D trải qua 2 bước:

A. Enrollment: Dữ liệu thô được chuyển đổi thành metadata và lưu trữ trong cơ sở dữ liệu qua các bước sau:

- **Acquisition:** dữ liệu thô được thu thập từ cảm biến và được biểu diễn dưới dạng đa giác
- **Registration:** Dữ liệu được căn chỉnh vào một hệ tọa độ hợp nhất bằng cách kết hợp ba thuật toán căn chỉnh khác nhau
- **Deformable Model Fitting:** AFM được fit dựa vào một deformable model framework
- **Geometry Image Analysis:** từ AFM thu được geometry map và normal map, đồng thời dùng phân tích wavelet để trích xuất metadata



Quy trình enrollment

B. Authentication:

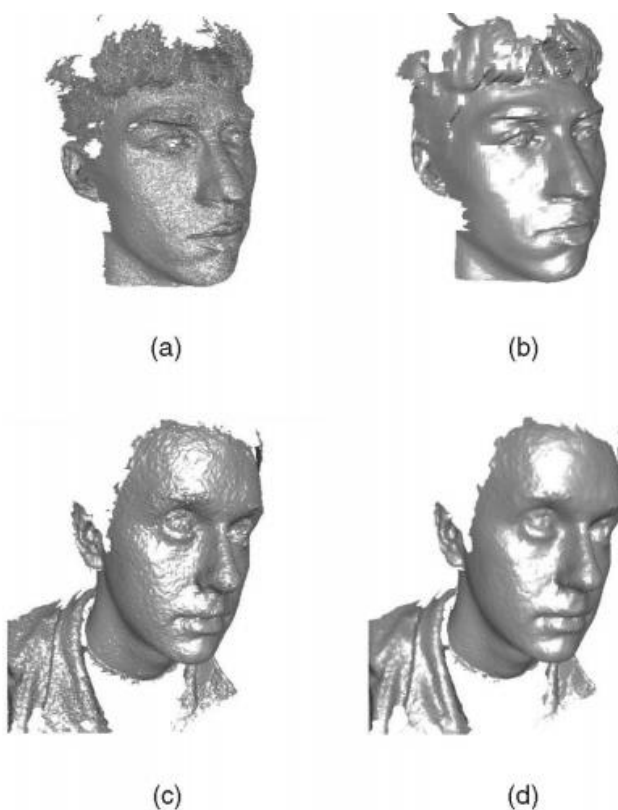
Metadata truy xuất từ cơ sở dữ liệu được dùng để so sánh đối chiếu bằng một số độ đo khoảng cách.

1.1. Tiền xử lý dữ liệu:

Mục đích: Loại bỏ một số vấn đề về cảm biến, đồng thời thống nhất dữ liệu từ các cảm biến khác nhau

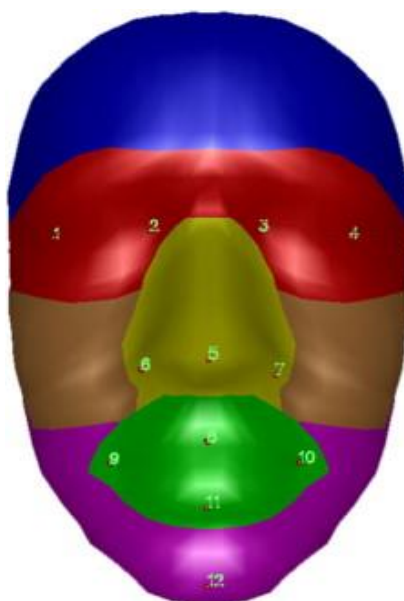
Bước tiền xử lý bao gồm các bộ lọc được thực hiện theo thứ tự như sau:

- **Median Cut:** được dùng để khử gai cho dữ liệu. Vấn đề này thường xuất hiện phổ biến hơn trong camera laser.
- **Hole filling:** các camera laser thường tạo ra các lỗ hổng trên bề mặt (vị trí mắt, lông mày,...) do đó cần áp dụng hole filling để lấp các lỗ hổng này
- **Smoothing:** khử nhiễu trắng từ các máy quét độ phân giải cao
- **Subsampling:** Bước fit deformable model (mục 1.1.4) sẽ resample dữ liệu, làm cho phương pháp này không nhạy cảm với độ phân giải dữ liệu. Do đó, ta giảm độ phân giải đến một mức nhất định để vừa đạt được hiệu quả về mặt lưu trữ, vừa không làm giảm độ chính xác



1.2 Annotated Face Model (AFM)

Sử dụng mô hình chú thích mặt người (AFM), gồm có 12 landmark và 6 vùng: trán, mắt, mũi, miệng, má và cằm, được thể hiện như hình bên dưới



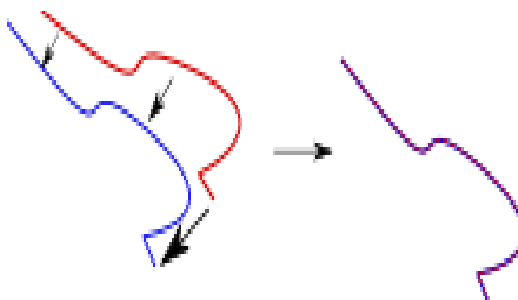
Mô hình này được sử dụng trong bước alignment, fitting và rút trích metadata

1.3. Registration (Pose correction - căn chỉnh góc mặt):

Registration là quá trình biến đổi các bộ dữ liệu khác nhau về cùng một hệ tọa độ.

Đây là một bước quan trọng đối với bất kỳ phương pháp hình học nào. Trong thực tế, nếu xảy ra lỗi trong bước căn chỉnh này thì sẽ ảnh hưởng lớn đến các bước phía sau. Sau đây là quy trình căn chỉnh đa tầng, gồm có 2 bước:

- **Iterative Closest Point (ICP):** là thuật toán dùng để cực tiểu hóa độ khác biệt giữa 2 tập điểm

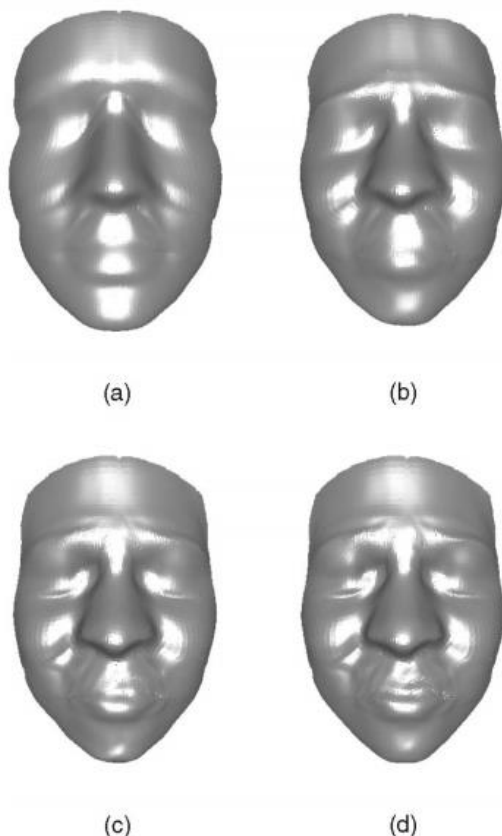


Tư tưởng chính của ICP

- **Simulated Annealing on Z-Buffers:**
 - Simulated Annealing (thuật toán mô phỏng luyện kim)
 - Z-buffer: còn được gọi là depth-buffer là 1 mảng 2 chiều với chỉ số là tọa độ x, y và giá trị là tọa độ z
 - Ý tưởng chính của bước này là giảm thiểu sự khác biệt giữa z-buffer của mô hình và dữ liệu.

Quy trình được đề xuất là kết quả của việc thực nghiệm trên bộ dữ liệu khuôn mặt. Bên cạnh đó, quy trình này cũng đạt hiệu quả trong trường hợp tổng quát.

1.4. Deformable Model Fitting



Tiến trình fit: AFM sau a-0 b-8 c-32 d-64 lần lặp

Mô hình AFM được fit cho từng bộ dữ liệu riêng lẻ để thu được những đặc điểm về hình học trên khuôn mặt. Điều này đạt được bằng cách sử dụng Deformable model framework của Metaxas and Kakadiaris, kết hợp với giải pháp Loop subdivision surfaces của Mandal.

1.5. Phân tích ảnh Geometry:

Mô hình thu được từ bước fitting sẽ được chuyển đổi thành ảnh Geometry, bằng cách lấy mẫu trên bề mặt của mô hình và mã hoá thông tin lên lưới 2D. Độ phân giải của lưới này phụ thuộc vào độ phân giải của bề mặt phân khu trên mô hình AFM.

Sau đó, từ ảnh Geometry ta xây dựng normal map. Trái với Geometry chỉ tập trung vào thành phần Z, Normal map phân phối thông tin đồng đều giữa 3 thành phần X, Y, Z.



Ảnh Geometry



Normal map

Cuối cùng, sử dụng phép biến đổi Haar và Pyramid để thu được hai tập hệ số, sử dụng làm metadata. Cụ thể, ta áp dụng phép Haar trên cả ảnh geometry và normal, trong khi đó chỉ sử dụng phép Pyramid trên ảnh geometry.

1.6. Các độ đo cho bước xác thực:

Trong bước xác thực, phép so sánh hai đối tượng trong gallery và probe sẽ được thực hiện trên metadata. Nhóm tác giả giới thiệu một phương pháp tiếp cận mới, sử dụng kết hợp cả hai độ đo Haar và Pyramid (tương ứng với hai phép biến đổi ở trên)

- **Độ đo Haar:**

Sử dụng độ đo Manhattan trên từng thành phần, ví dụ với X:

$$d_x^h(P, G) = \sum_{i,j} |P_x[i, j] - G_x[i, j]|,$$

Trong đó P, G là ảnh probe, gallery
Tổng khoảng cách giữa P và G là tổng trên từng thành phần:

$$d^h(P, G) = d_x^h(P, G) + d_y^h(P, G) + d_z^h(P, G).$$

- **Độ đo Pyramid:** Sử dụng thuật toán CW-SSIM, dùng một cửa sổ trượt để so sánh từng phần của ảnh.

TABLE 1
Verification Rates of Our System at 0.001 FAR Using Different Transforms on the FRGC v2 Database

	ROC I	ROC II	ROC III
Fusion	97.3%	97.2%	97.0%
Haar	97.1%	96.8%	96.7%
Pyramid	95.2%	94.7%	94.1%

Hiệu quả khi sử dụng độ đo Haar, Pyramid và kết hợp cả hai

2. 3D Face Recognition for Partial Scans: UR3D-PS:

Phương pháp UR3D chỉ tập trung vào việc quét 3D mặt trước khuôn mặt và không xử lý được khi dữ liệu bị thiếu. Trong phần này ta sẽ giải quyết vấn đề này bằng việc tiến hành quét 3D một phần khuôn mặt và xử lý trường hợp bị thiếu dữ liệu (ví dụ như quét một bên mặt với góc quay lớn).

Giải pháp được đặt ra là kết hợp các lần quét khuôn mặt một cách xen kẽ và sử dụng tính chất đối xứng của khuôn mặt (sẽ nói đến ở dưới).

Mục tiêu của phần này là kết hợp giữa việc quét cả mặt trước và mặt bên và sử dụng kết quả quét để tạo ra được 1 chữ ký sinh trắc học bất biến, do đó phù hợp với việc áp dụng trong các trường hợp thực tế (luôn luôn thay đổi và có sự tác động của điều kiện môi trường).

Trong phương pháp này việc quét mặt trước thường được thực hiện một cách độc lập đối với từng bên của khuôn mặt, kết hợp với việc sử dụng phép đối xứng và sau đó so khớp với nhau. (áp dụng tính bất biến của AFM từ đó làm giảm bớt việc bị thiếu dữ liệu).

2.1 3D Landmark Detection:

Phương pháp UR3D-PS sử dụng phiên bản cải tiến của thuật toán **3D landmark detection**. Các landmarks ứng viên được xác định và gắn nhãn thông qua nhiều lần quét, áp dụng **Facial Landmark Model (FLM)**. 8 landmark được

sử dụng (FLM8): 2 góc trong của mắt và 2 góc ngoài của mắt, chóp mũi, góc phải, góc trái miệng và chóp cằm

Note: luôn luôn có ít nhất 5 điểm được nhìn thấy trong mỗi lần quét một bên của khuôn mặt. (FLM5R, FLM5L).

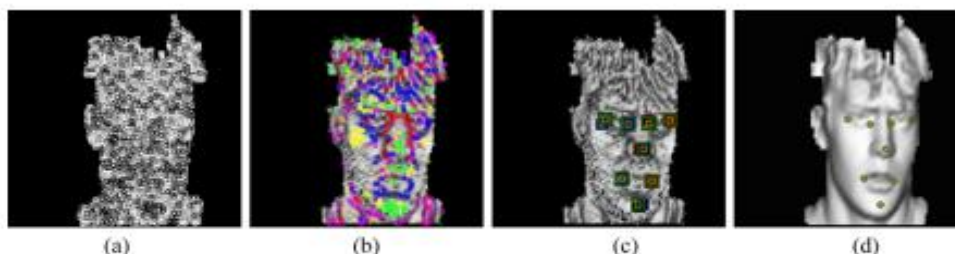


Fig. 17.5 Results of landmark detection and selection process: **a** shape indexes maxima and minima; **b** spin image classification; **c** extracted best landmark sets; and **d** resulting landmarks

- **Trích xuất landmarks ứng viên bằng cách sử dụng Shape Index map** (Extract candidate landmarks by using the Shape Index map): Sau khi tính được các chỉ số **shape index** bằng cách quét 3D khuôn mặt, ta ánh xạ các giá trị này vào không gian 2D từ đó xây dựng **Shape Index map**.
- **Shape Index map** bao gồm Caps (cực đại địa phương như chóp mũi, chóp cằm,...) và Cups (cực tiểu địa phương như góc mắt, góc miệng,...). Tập hợp con bao gồm các điểm quan trọng nhất của mỗi tập (Caps và Cups) được giữ lại.
- **Phân lớp landmarks ứng viên sử dụng Spin Image templates** (Classify candidate landmarks by using Spin Image templates): các landmarks ứng viên ở bước trước được phân loại và trích lọc theo mức độ phù hợp với các mẫu Spin Image có sẵn. Giả sử có 2 Spin Image là P và Q thì mối quan hệ của P và Q được thể hiện bằng cách so khớp với nhau và được biểu diễn trực quan bằng hệ số tương quan tuyến tính chuẩn hóa:

$$S(P, Q) = \frac{N \sum p_i q_i - \sum p_i \sum q_i}{\sqrt{[N \sum p_i^2 - (\sum p_i)^2][N \sum q_i^2 - (\sum q_i)^2]}}$$

○

Trong đó p_i và q_i là giá trị của các phần tử trong N phần tử của Spin Image P và Q tương ứng.

- **Gắn nhãn Landmarks** (Label Landmarks): Từ các landmarks ứng viên được phân loại, ta kết hợp 5 landmarks khả thi (có thể kết hợp) với nhau sau đó so khớp với các landmarks có sẵn. Nếu kết quả phù hợp với mô hình FLM5L hoặc FLM5R thì sẽ được dán nhãn, phân loại vào mô hình FLM tương ứng và được coi là 1 phương án kết hợp khả thi còn không thì sẽ bị loại ra.

Lưu ý: các phương án khả thi có thể bao gồm cả sự kết hợp giữa FLM5L và FLM5R sao cho phù hợp với mô hình FLM8.

- **Chọn kết quả cuối cùng** (Select Final Solution): kết quả tối ưu cho FLM5L, FLM5R, FLM8 được chọn ra dựa trên sự sai khớp khi so với FLM tương ứng. Để chọn ra kết quả cuối cùng là sự kết hợp giữa 3 mô hình ta dùng **Procrustes Distance** (có tính đến kích thước hình dạng không gian).

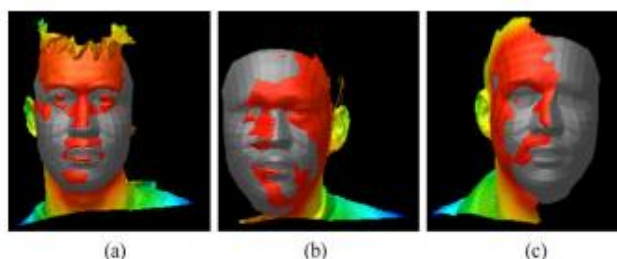


Fig. 17.6 AFM (gray) and facial scans (color coding: red means low registration error, blue means high registration error) superposed after registration (the scans): **a** frontal scan; **b** 45° left side scan; and **c** 60° right side scan

2.2 Symmetric Deformable Model Fitting:

Framework có thể xử lý một cách độc lập mặt trái và phải của AFM nhờ vào sử dụng tính đối xứng của khuôn mặt.

Ý tưởng: nhờ vào tính đối xứng của khuôn mặt để tránh việc tính toán trên những vùng bị thiếu dữ liệu. Ví dụ: khi mặt bên phải bị thiếu dữ liệu ta áp dụng AFM vào mặt bên trái và chỉ việc lấy đối xứng kết quả sẽ suy ra được mặt còn lại.

3. 3D-aided 2D Face Recognition: UR2D:

UR2D là phương pháp Face Recognition hỗ trợ cả dữ liệu 2D và 3D

3.1 3D + 2D Enrollment:

UR2D Enrollment sử dụng **Annotated Face Model (AFM)** được đề xuất bởi Kakadiaris để tạo ra những ảnh dạng hình học (dùng 3 channels để cập nhật liên tục những hình ảnh 2D) sau đó mã hóa các thông tin hình học thu được (cụ thể là các thành phần x,y,z trong không gian 3 chiều).

- 7 channels dùng cho ảnh dạng hình học
 - + 3 channels để biểu diễn cấu trúc hình học thực tế của khuôn mặt
 - + 3 channels để biểu diễn thông tin kết cấu khuôn mặt
 - + 1 channel dùng cho visibility map

Trong thực tế, ta thường sử dụng độ phân giải 256x256. Đầu tiên ta điều chỉnh AFM phù hợp với dữ liệu 3D đầu vào và biểu diễn dưới dạng geometry images. Với mỗi đỉnh của geometry images, ta sẽ tính điểm gần nhất trên dữ liệu. Texel tương ứng với điểm này trong data sẽ được sử dụng để tạo hình ảnh kết cấu ứng với AFM đã được điều chỉnh. Ngoài ra, visibility map sẽ được tính toán. Nếu điểm gần nhất với dữ liệu không được gán texel hợp lệ (nghĩa là những điểm 3D này không phản ứng với cảm biến ảnh 2D), nó sẽ được gán cho vị trí tương ứng trên visibility map ngược lại nó được gán bằng 0.

Thuật toán: Enrollment với dữ liệu 3D

Input: 3D facial mesh, subject ID, ảnh khuôn mặt dưới dạng 2D

1. Tiền xử lý 3D facial mesh

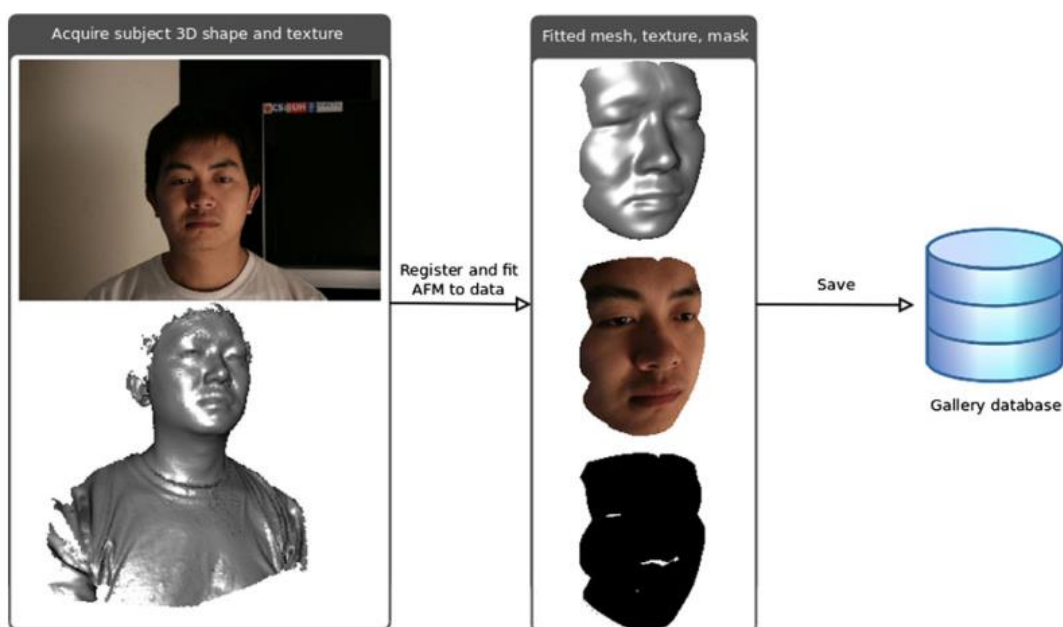
2. Áp dụng AFM với 3D facial mesh
3. Điều chỉnh AFM sao cho phù hợp với 3D facial mesh
4. Dựa vào AFM để biến đổi kết cấu từ ảnh input 2D
5. Tính Visibility map
6. Lưu AFM đã được điều chỉnh, kết cấu ảnh đã được biến đổi và visibility map trong bộ dữ liệu thu thập ứng với subject ID

Table 17.1 Variations of the UR2D system

Method name	Gallery Data	Probe Data	Geometry Image	Relighted	Distance metric	Score Normalization
UR3D	3D	3D	X		CWSSIM	MAD
UR2D-V-1	3D + 2D	2D	X	X	GS	E
UR2D-V-2	3D + 2D	2D	X	X	CWSSIM	E
UR2D-V-3	3D + 2D	2D	X		CWSSIM	E
L1	2D	2D				
UR2D-V-4	2D	2D			GS	E

3.2 2D Authentication:

Trong giai đoạn Authentication, đầu vào là hình ảnh 2D, the seven landmark (2 góc trong của mắt, 2 góc ngoài của mắt, chóp mũi và 2 góc mũi) được phát hiện bằng PittPatt. Khi góc mặt được ước tính (sử dụng các mốc và các vị trí tương ứng trên AFM), kết cấu sẽ được ánh xạ lên AFM.



Thuật toán: Authentication với dữ liệu 2D

Input: ảnh khuôn mặt dưới dạng 2D, claimed subject ID

1. Truy xuất “claimed ID” AFM từ gallery
2. Xác định vị trí “the seven landmark” trên ảnh khuôn mặt 2D

3. Đặt AFM vào hình ảnh mặt 2D ở các mốc tương ứng
4. Tính visibility map
5. Điều chỉnh 2 chiều kết cấu khuôn mặt 2D enrollment để phù hợp với kết cấu khuôn mặt 2D probe
6. Tính CWSSIM và GS scores giữa relit texture và probe texture
7. Đặt ra ngưỡng của score để đưa ra quyết định ACCEPT/REJECT

3.3 Skin Reflectance Model :

Trong trường hợp dữ liệu thực nghiệm có đủ độ phân giải, có một hàm phân phối (BSSRDF) dùng để mô hình hóa phản xạ từ da. Tuy nhiên trong hầu hết các hệ thống nhận dạng, dữ liệu có độ phân giải khá thấp

IV. Thực nghiệm:

Ở đây ta chỉ phân tích kết quả của phương pháp UR3D và UR3D-PS

1. 3D Face Recognition: UR3D

Đối với phương pháp này ta sử dụng cơ sở dữ liệu FRGC v2 bao gồm 4007 ảnh của 466 người thu được bằng cách quét 3D mặt trước của khuôn mặt.



Fig. 17.14 Facial scans with various expressions for a subject from the FRGC v2 database

Để tính toán hiệu suất của phương pháp UR3D trên bộ dữ liệu này, ta chuẩn bị 3 bộ test với độ khó tăng dần được gọi lần lượt là **ROC I**, **ROC II** và **ROC III**.

The verifications rates của phương pháp được cho ở mức 0.001 **False Acceptance Rate (FAR)**

Table 17.2 Verification rates of our method at 0.001 FAR using different transforms on the FRGC v2 database

	ROC I	ROC II	ROC III
Fusion	97.3%	97.2%	97.0%
Haar	97.1%	96.8%	96.7%
Pyramid	95.2%	94.7%	94.1%

Kết quả được trình bày trực quan thông qua biểu đồ:

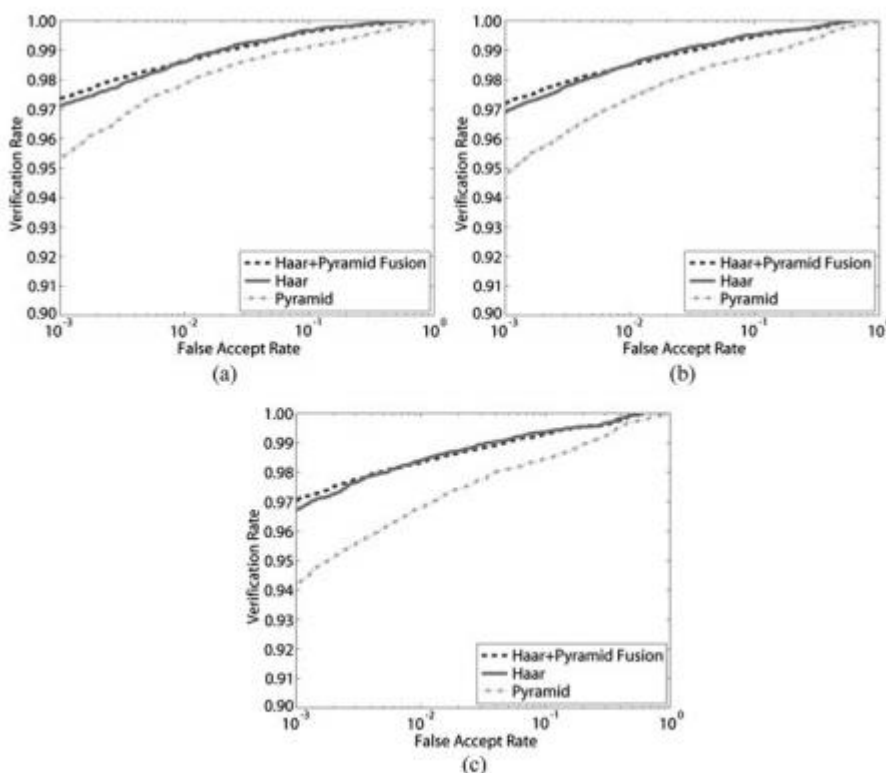


Fig. 17.15 Performance of the proposed method using the Haar and Pyramid transforms as well as their fusion on the FRGC v2 database. Results reported using: **a** ROC I, **b** ROC II, and **c** ROC III

Giá trị trung bình của The verifications rates trên ROC I, ROC II và ROC III là 97.16% đối với phương pháp kết hợp giữa Haar và Pyramid. Giá trị này rơi vào khoảng 96.86% đối với phép biến đổi Haar và 94.66% đối với phép biến đổi Pyramid. Từ đây ta có thể nhận thấy Pyramid tốn kém hơn trong việc tính toán và thể hiện ra sự vượt trội của Haar. Sự kết hợp của 2 phép biến đổi cho ra kết quả tốt hơn đặc biệt là đối với các bộ test khó (ROC II và ROC III).

2. 3D Face Recognition for Parital Scans: UR3D-PS

Cơ sở dữ liệu dùng trong phương pháp này là sự kết hợp giữa FRGC v2 và UND (trong đó FRGC v2 bao gồm dữ liệu thu được nhờ quét mặt trước của khuôn mặt và UND là quét mặt bên).

UND là bộ dữ liệu dùng trong việc nhận dạng tai bao gồm các hình ảnh quét bên trái và bên phải của khuôn mặt với các góc quét 45, 60 và 90 độ. Trong trường hợp dùng bộ dữ liệu này để chạy thực nghiệm phương pháp, ta sẽ ngầm hiểu bộ dữ liệu này sẽ thể hiện cho trường hợp ta quét một phần khuôn mặt và bị thiếu dữ liệu.

Ta sẽ tạo ra 2 tập dữ liệu lần lượt gọi là UND45LR và UND60LR tương ứng với các góc quét 45 và 60 độ trong đó UND45LR bao gồm 118 đối tượng và

UND60LR là 87 đối tượng. Trong đó mỗi đối tượng ứng với 2 trường hợp quét bên trái và bên phải (bên trái được gọi là gallery và bên phải gọi là probe).

Fig. 17.16 Left and right side facial scans from the UND Ear Database



Ngoài ra ta còn có 1 tập dữ liệu được gọi là UND00LR được định nghĩa bao gồm 2 phần: gallery và probe. Trong đó gallery thu được nhờ quét mặt trước của 466 đối tượng trong cơ sở dữ liệu FRGC v2 và probe thu được nhờ quét 45 độ cả trái lẫn phải của 39 đối tượng và quét 60 độ cả trái lẫn phải của 32 đối tượng. Chỉ những đối tượng ở trong tập gallery mới được đưa vào probe.

Table 17.3 Rank-one recognition rate of our method for matching partial scans

	Rank-one Rate
UND45LR	86.4%
UND60LR	81.6%
UND00LR	76.8%

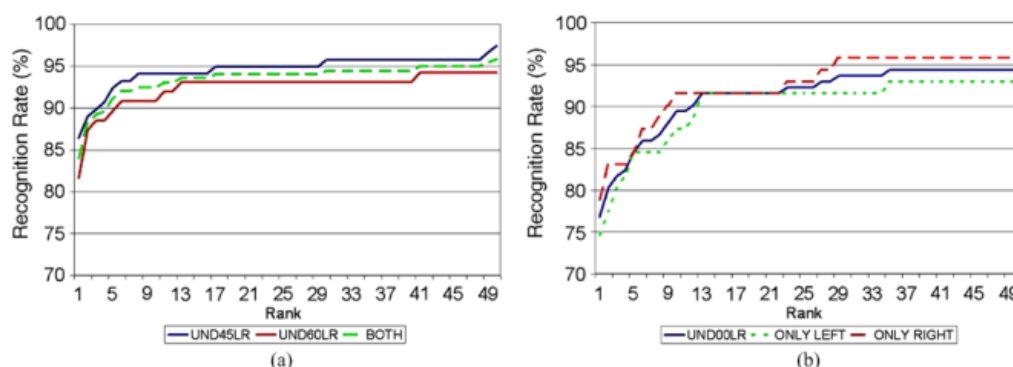


Fig. 17.17 **a** CMC graphs for matching left (gallery) with right (probe) side scans using UND45LR, UND60LR and the combination of the two; **b** CMC graphs for matching frontal (gallery) with left, right and both (probe) side scans using UND00LR

Dựa theo kết quả ta có thể thấy phương pháp cho kết quả tốt hơn đối với tập dữ liệu UND45LR và UND60LR. Mặc dù tập dữ liệu UND00LR bao gồm cả hình quét trực diện không bị thiếu dữ liệu sẽ cho kết quả tốt hơn nhưng do sự chênh lệch về kích thước so với UND45LR và UND60LR từ đó dẫn đến nhiều khó khăn hơn trong quá trình thử nghiệm.

Tập dữ liệu UND60LR với góc quét lớn hơn cho kết quả không tốt bằng UND45LR.

V. Demo:

Demo thực hiện việc nhận dạng khuôn mặt sử dụng thư viện PyTorch.

- **Dataset:** tập dữ liệu gồm 403067 bộ hình ảnh của 1208 người. Mỗi bộ hình ảnh sẽ gồm 1 ảnh RGB và một ảnh chiều sâu.
- **Mô hình được sử dụng trong demo:** *ResNet50* sử dụng cho việc rút trích đặc trưng, với kênh input được điều chỉnh một chút cho phù hợp với dữ liệu và sử dụng hàm softmax.

ResNet50 là một mạng convolutional neural network sâu 50 tầng. *ResNet50* có thể sử dụng dữ liệu huấn luyện trước, được huấn luyện từ hàng triệu tấm ảnh trên cơ sở dữ liệu ImageNet. Mạng huấn luyện trước khả năng phân lớp 1000 loại đối tượng và học nhiều đặc trưng với nhiều loại hình ảnh. Mạng ***ResNet50*** có kích thước ảnh đầu vào là 224x224.

- **Bắt đầu với giai đoạn tiền xử lý dữ liệu:** do *ResNet50* cần có dữ liệu đầu vào là hình ảnh với kích thước cố định, do đó ta tiến hành face detection & alignment bằng Multi-task CNN. Sau đó, tập dữ liệu sẽ được chia thành 3 tập dữ liệu huấn luyện, kiểm tra và đánh giá. Thông tin về các tập dữ liệu sau khi chia được lưu vào file csv.
- **Quá trình huấn luyện:** Tập dữ liệu huấn luyện và đánh giá được sắp xếp và đưa vào input. Việc huấn luyện được thực hiện trên một số lượng epochs nhất định. Với mỗi epoch:
 - Sử dụng *ResNet50*, tiến hành huấn luyện trên tập dữ liệu, cập nhật SGD và cross validation.
 - Tiến hành đánh giá trên tập đánh giá để tính toán độ chính xác và mất mát của epoch vừa tiến hành huấn luyện và so sánh với độ chính xác cao nhất.

Từ đó tìm ra được mô hình có độ chính xác cao nhất và sử dụng mô hình đó cho quá trình nhận dạng

Các bước tiến hành demo: (<https://github.com/XingwXiong/Face3D-Pytorch>)

(Demo được chạy trên hệ điều hành Linux)

B1: Clone repository về máy

B2: Cài đặt các thư viện cần thiết cho demo bằng cách chạy câu lệnh trên

Terminal:

```
pip install -r requirements
```

B3: Tiến hành download tập dữ liệu

(http://125.39.136.212:8484/3dvggface2_1.tar.gz) và giải nén.

B4: Thực hiện Face alignment bằng cách chạy câu lệnh trên Terminal:

```
python preprocess/align/align_dataset_mtcnn.py \
```

```
--input_dir '~/vggface3d' \
```

```
--output_dir '~/vggface3_align' \
```

```
--image_size 182 \
```

```
--margin 44 \  
--random_order \  
--thread_num 3 \  
--gpu_memory_fraction 0.88
```

B5: Thực hiện chia tập dữ liệu thành các tập dữ liệu con bằng cách chạy câu lệnh:

```
python preprocess/get_dataset_csv.py
```

B6: Thực hiện việc huấn luyện bằng cách chạy câu lệnh:

```
python train_softmax.py --train_dataset_csv '~/vggface3d_sm/train.csv' \  
--eval_dataset_csv '~/vggface3d_sm/eval.csv' \  
--pretrained_on_imagenet \  
--input_channels 4 \  
--num_of_classes 1200 \  
--num_of_epochs 50 \  
--num_of_workers 8 \  
--log_base_dir './logs'
```

B7: Thực hiện việc đánh giá bằng cách chạy câu lệnh:

```
python evaluation.py \  
--test_dataset_csv '~/vggface3d_sm/test.csv' \  
--pretrained_model_path './RGB-D-ResNet50-from-scratch.pkl' \  
--num_of_workers 8
```

VI. State of the art:

1.. Method:

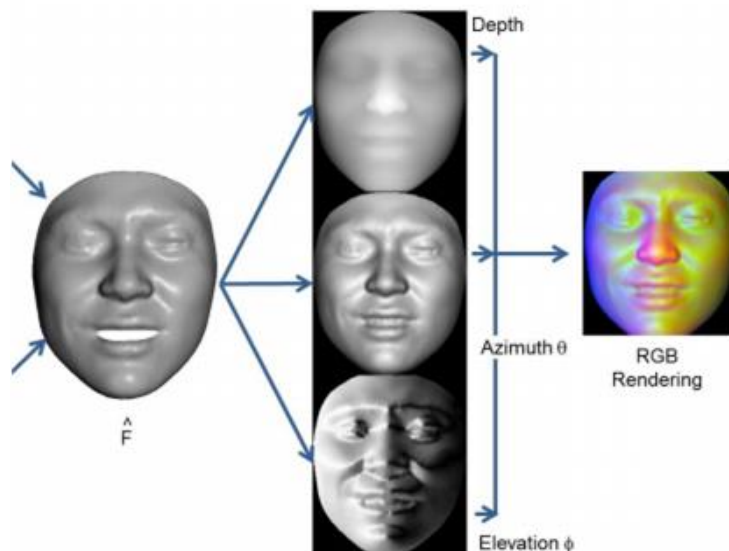
Ở đây, nhóm sẽ trình bày các nội dung chính của phương pháp FR3DNet (state-of-the-art) và trình bày tư tưởng của K3DM

1.1 FR3DNet:

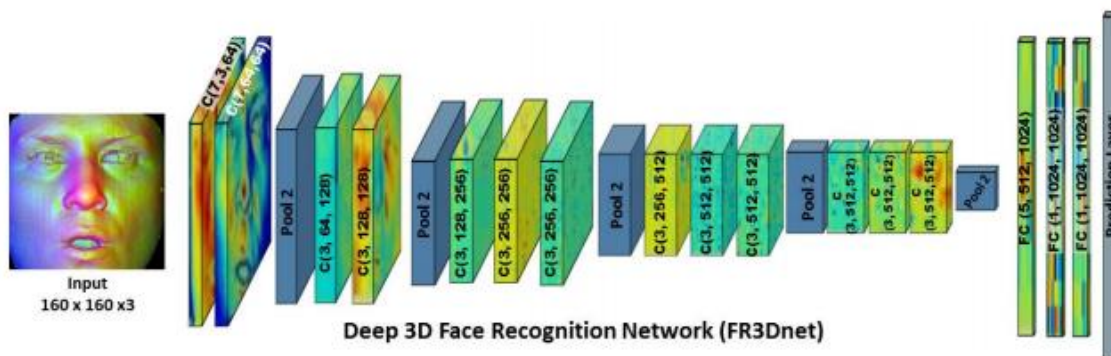
- **Sinh tập dữ liệu huấn luyện:**

Dữ liệu 3D point-cloud được chuyển đổi thành ảnh 3 channels sử dụng thuật toán gridfit. 3 channels này sau đó được chuẩn hoá về khoảng [0, 255] và có thể xem như ảnh RGB.

Ảnh này sau đó được đưa qua một mạng để xác định vị trí đầu mũi và tiến hành crop ảnh về kích thước 224 x 224 với đầu mũi nằm ở trung tâm. Cuối cùng ảnh được down-sampled về kích thước 160 x 160 để sử dụng huấn luyện.



- Kiến trúc mạng FR3Dnet:



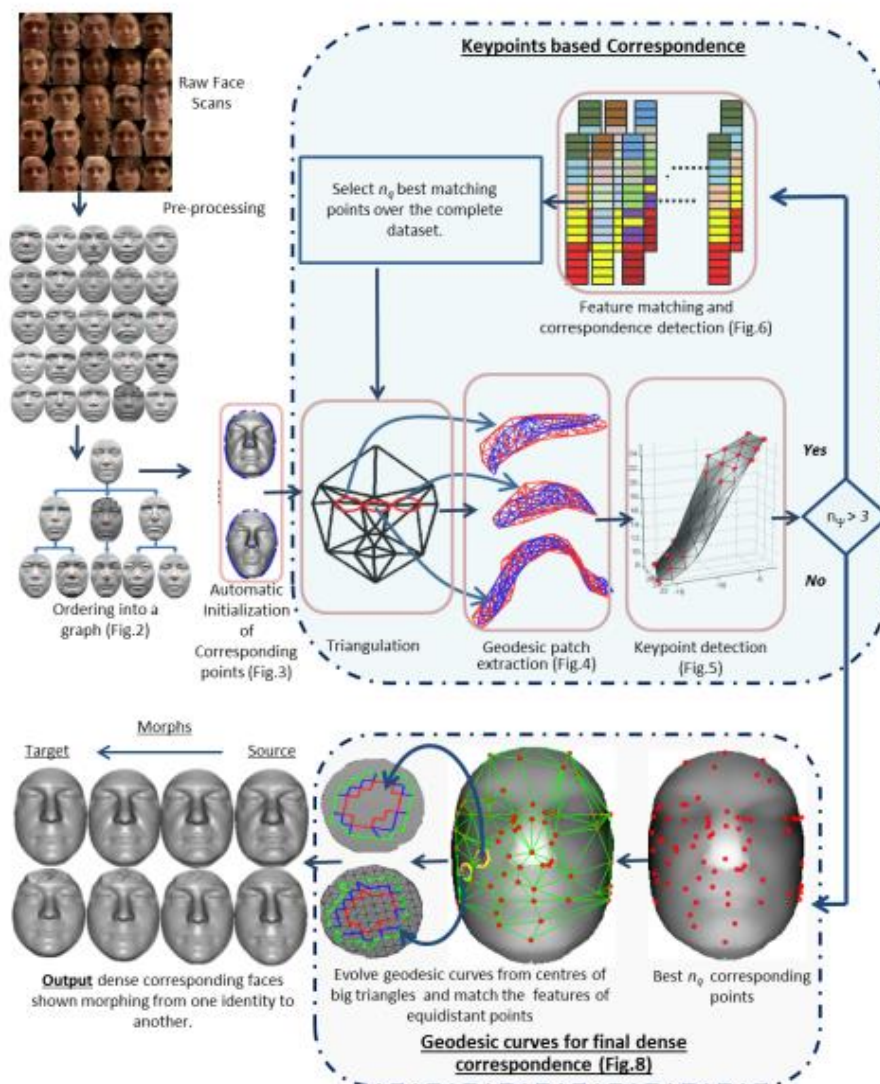
Kiến trúc mạng FR3DNet được trình bày như ở hình trên, với đầu vào có kích thước $160 \times 160 \times 3$ (channels) và mỗi convolutional layer được theo sau bởi một rectifier layer. Để đạt kết quả tốt hơn, nhóm nghiên cứu đã thiết kế mạng cho việc học $N = 100,005$ đối tượng riêng biệt.

Sau khi mạng được huấn luyện, tiến hành loại bỏ các dropout layer. Vector đặc trưng (có kích thước 1024) từ FC layer được sử dụng cho việc nhận dạng bằng cách cực tiểu hoá khoảng cách cosine giữa probe và gallery trong không gian đặc trưng.

Nhóm nghiên cứu cũng tiến hành fine-tune FR3DNet trên tập gallery (ký hiệu là FR3DnetFT) để đạt được độ chính xác cao hơn.

1.2 K3DM:

Tác vụ chính của phương pháp này là tìm sự tương đồng giữa một tập rất lớn các hình dạng tương tự nhau, ở đây là mặt người 3D. Mục đích của tác vụ này là dựa vào sự tương đồng, ta có thể tạo ra được một 3D morphable models với độ chính xác cao, sử dụng cho việc chuyển đổi dữ liệu từ tập huấn luyện và ảnh test. Bên dưới là sơ đồ thuật toán cho tác vụ này.



Sơ đồ khối của thuật toán Dense 3D face correspondence

2. Performance:

Method	Model \ Technique	Modality	Gallery of LS3DFace										
			LS3DFace This paper	FRGC [49]	BU3DFE [65]	BU4DFE [64]	Bosphorus [51]	CASIA [63]	GavabDB [41]	TexasFRD [23]	3D-TEC [61]	UMBDDB [15]	ND-2006 [19]
CNN	GoogleNet [57]	RGB	53.97	21.51	50.76	65.41	63.44	85.91	-	53.08	79.95	65.78	24.14
	Resnet152 [24]	RGB	15.05	13.53	8.04	9.64	7.05	52.85	-	20.94	72.66	34.08	10.92
	VGG-Face [45]	RGB	90.85	87.92	97.68	96.51	96.39	94.18	-	99.73	83.30	81.54	82.86
	GoogleNet [57]	3D	38.66	35.54	46.56	41.88	26.81	50.81	66.56	67.59	67.29	47.66	30.81
	Resnet152 [24]	3D	12.49	14.40	5.80	10.13	3.84	25.34	44.26	16.25	60.98	22.20	12.08
	VGG-Face [45]	3D	61.20	62.42	71.16	53.17	48.14	71.95	77.38	85.58	78.04	67.48	60.81
Conventional	MMH [35]	3D + 2D	83.08	89.37	88.50	84.93	85.10	85.24	86.64	85.67	80.85	77.32	86.71
	3D Keypoint [36]	3D	81.76	86.59	85.14	82.50	82.64	81.38	84.41	84.99	75.63	71.68	82.30
	R3DM [9]	3D	82.89	87.50	87.13	83.21	86.06	84.51	85.60	85.47	78.27	77.11	84.84
	K3DM [13]	3D	84.67	89.50	89.24	86.05	88.60	85.35	87.90	86.13	79.55	78.64	87.77
CNN	FR3DNet	3D	95.51	97.06	98.64	95.53	96.18	98.37	96.39	100.00	97.90	91.17	95.62
	FR3DNet _{FT}	3D	98.75	99.88	99.96	98.04	100.00	99.74	99.70	100.00	99.12	97.20	99.13

Bảng so sánh độ chính xác của một số phương pháp sử dụng mạng học sâu và phương pháp truyền thống

Xét mảng 3D, ta thấy **FR3DNet** đạt được độ chính xác cao nhất, và cách khá xa các phương pháp khác trong tất cả các bộ dữ liệu (độ chính xác còn được tăng lên với mô hình đã fine tune)

Tiếp theo là **K3DM** (phương pháp truyền thống) tuy độ chính xác thấp hơn **FR3DNet** nhưng cũng khá cao so với các phương pháp còn lại.

VII. REFERENCES

- [1] C. Mandal, H. Qin, and B. Vemuri, "Dynamic Smooth Subdivision Surfaces for Data Visualization," Visualization, pp. 371-377, Oct. 1997.
- [2] C. Mandal, H. Qin, and B. Vemuri, "A Novel FEM-Based Dynamic Framework for Subdivision Surfaces," Computer-Aided Design, vol. 32, nos. 8-9, pp. 479-497, 2000.
- [3] C. Loop, "Smooth Subdivision Surfaces Based on Triangles," master's thesis, Dept. of Math., Univ. of Utah, 1987.
- [4] Kakadiaris, I., Passalis, G., Toderici, G., Murtuza, M., Lu, Y., Karampatziakis, N., Theoharis, T.: Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. IEEE Trans. Pattern Anal. Mach. Intell. 29(4), 640-649 (2007)
- [5] Syed Zulqarnain, and Gilani Ajmal Mian, "Learning from Millions of 3D Scans for Large-scale 3D Face Recognition", CVPR 2018
- [6] Syed Zulqarnain Gilani, Ajmal Mian, Faisal Shafait, and Ian Reid, "Dense 3D Face Correspondence"

HẾT